



Технології BigData

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	<i>Другий (магістерський)</i>
Галузь знань	<i>F Інформаційні технології</i>
Спеціальність	<i>F7 Комп'ютерна інженерія</i>
Освітня програма	<i>Комп'ютерні системи та мережі</i>
Статус дисципліни	<i>Нормативна</i>
Форма навчання	<i>очна(денна)</i>
Рік підготовки, семестр	<i>1 курс, осінній семестр</i>
Обсяг дисципліни	<i>4 кредитів, лекції 30 годин, лабораторні 14 годин, СРС 76 годин.</i>
Семестровий контроль/ контрольні заходи	<i>Залік</i>
Розклад занять	<i>Згідно розкладу на весняний семестр поточного навчального року за адресою http://rozklad.kpi.ua/</i>
Мова викладання	<i>Українська</i>
Інформація про керівника курсу / викладачів	Лектор: <i>ст. викладач, доктор філософії, Таран Владислав Ігорович, taran@comsys.kpi.ua</i> Лабораторні: <i>ст. викладач, доктор філософії, Таран Владислав Ігорович, taran@comsys.kpi.ua</i>
Розміщення курсу	https://classroom.google.com/c/Nzg4MTMwNjY4MTk1 https://cloud.comsys.kpi.ua/s/weFE5o9BwpjEs76

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Дисципліна “Технології BigData” спрямована на вивчення підходів, методів і механізмів функціонування та використання інфраструктури для розподілених обчислень на базі кластеру Hadoop та парадигми MapReduce. Необхідність в використанні нових підходів обумовлена тим, що сучасні підходи до вирішення складних завдань, які потребують обробки надзвичайно великого обсягу даних, потребують використання великої кількості обчислювальних ресурсів. Вивчення даної дисципліни майбутніми фахівцями дозволить їм набути важливих компетенцій в плані розвитку існуючих і використанню нових підходів для організації розподілених обчислень.

Метою вивчення дисципліни “Технології BigData” є підготовка фахівців, здатних розв'язувати комплексні проблеми у сфері обробки “великих даних” та використовувати сучасні засоби для організації обчислень в розподілених системах.

Предметом дисципліни є:

- підходи та методи обробки “великих даних” в розподілених системах;
- механізми організації обчислень в розподілених системах.

Дисципліна “Технології BigData” забезпечує наступні програмні компетентності і програмні результати освітньо-наукової програми (ОНП): ФК7, ФК15, ПРН1, ПРН7, ПРН15.

Згідно з вимогами ОНП здобувачі після засвоєння дисципліни “Технології BigData” мають продемонструвати такі програмні результати навчання:

- Здатність досліджувати, розробляти та обирати технології створення великих і надвеликих систем.;
- Здатність розробляти системи обробки великих обсягів даних;
- Застосовувати загальні підходи пізнання, методи математики, природничих та інженерних наук до розв’язання складних задач комп’ютерної інженерії;
- Вирішувати задачі аналізу та синтезу комп’ютерних систем та мереж;
- Здійснювати дослідження та проектування високопродуктивних комп’ютерних систем.

За результатами вивчення навчальної дисципліни “Технології BigData” мають бути отримані такі **знання**:

- Мати передові концептуальні та методологічні знання у сфері обробки та аналізу великих обсягів даних;
- Мати методологічні знання в плані застосування сучасних підходів та засобів для організації обчислень у розподілених обчислювальних системах.

Уміння, які мають бути отримані у рамках вивчення навчальної дисципліни “Технології BigData”:

- Вміти ефективно здійснювати пошук та критичний аналіз організації розподілених обчислювальних систем;
- Вміти розв’язувати задачі налаштування основних компонентів розподілених обчислювальних систем;
- Вміти розробляти програмне забезпечення для обробки великих даних у розподілених обчислювальних системах;
- Вміти застосовувати технології організації обчислень у розподілених системах;
- Вміти застосовувати технології організації складних послідовностей обчислень у розподілених системах.

Здобувачі наукового ступеня також мають бути здатні:

- Застосовувати прикладні бібліотеки та програмні системи, які використовуються при розробці основних компонентів розподілених обчислювальних систем.
- Володіти методами та технологіями програмування з використанням прикладних бібліотек та програмних систем, призначених для інтеграції основних компонентів розподілених обчислювальних систем.

Таке поєднання загальних та спеціальних компетентностей, теоретичних та практичних знань, умінь та здатностей сприяє підвищенню науково-практичного рівня здобувачів наукового ступеня магістра задля здійснення ними ефективних наукових досліджень.

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Необхідні дисципліни: “Програмування”, “Архітектура комп’ютерних систем”, “Комп’ютерні системи”, “Структури даних та алгоритми”, “Дослідження і проектування комп’ютерних систем”, “Алгоритми та методи обчислень”, “Дискретна математика”.

3. Зміст навчальної дисципліни

Перелік основних тем, що входять до програми вивчення дисципліни “Технології BigData”:

Розділ 1. Вступ до Big Data

Тема 1.1. Основні поняття Big Data

Тема 1.2. Характеристика Big Data

Тема 1.3. Приклади задач Big Data

Тема 1.4. Сучасні технології Big Data

Розділ 2. Парадигма MapReduce

Тема 2.1. MapReduce у розподілених обчисленнях

Тема 2.2. Аспекти роботи MapReduce

Тема 2.3. Основні етапи MapReduce

Тема 2.4. Приклади застосування

Розділ 3. Платформа для розподілених обчислень Hadoop

Тема 3.1. Характеристики кластеру Hadoop

Тема 3.2. Розподілена файлова система – HDFS

Тема 3.3. Менеджер ресурсів кластеру – YARN

Тема 3.4. Засоби моніторингу роботи кластеру Hadoop

Тема 3.5. Запуск обчислювальних задач на кластері Hadoop

Тема 3.6. Особливості інсталяція та налаштування кластеру Hadoop

Розділ 4. Застосунок Apache Pig

Тема 4.1. Основні поняття Apache Pig

Тема 4.2. Порівняння з MapReduce

Тема 4.3. Мова Pig Latin

Тема 4.4. Написання скриптів для кластеру Hadoop

Розділ 5. Застосунок Apache Hive

Тема 5.1. Основні поняття Apache Hive

Тема 5.2. Порівняння між Apache Pig та MapReduce

Тема 5.3. Спосіб організації даних у Apache Hive

Тема 5.4. Мова запитів HiveQL

Тема 5.5. Написання запитів для кластеру Hadoop

Розділ 6. Диспетчер робіт Apache Oozie

Тема 6.1. Основні поняття Apache Oozie

Тема 6.2. Потіки робіт у Apache Oozie

Тема 6.3. Особливості налаштування диспетчеру

Тема 6.4. Засоби моніторингу робіт

Тема 6.3. Створення та налаштування власного потоку робіт для кластеру Hadoop

4. Навчальні матеріали та ресурси

Базова:

1. Таран, В. І. Технології Big Data. Практикум [Електронний ресурс] : навчальний посібник для здобувачів ступеня магістра за освітньою програмою «Комп'ютерні системи та мережі» спеціальності 123 Комп'ютерна інженерія / В. І. Таран, Ю. Г. Гордієнко, С. Г. Стіренко ; КПІ ім. Ігоря Сікорського. – Електронні текстові дані (1 файл: 2,27 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2022. – 56 с. – Назва з екрана. URI <https://ela.kpi.ua/handle/123456789/50014>
2. Alex Holmes. Hadoop in Practice. – Manning Publications, 2012;

Додаткова:

1. Tom White. *Hadoop: The Definitive Guide: Storage and Analysis at Internet Scale.* – O'Reilly Media, 2010;
2. Donald Miner, Adam Shook. *MapReduce Design Patterns.* – O'Reilly Media, 2012.
3. Apache Hadoop (<https://hadoop.apache.org/docs/r2.10.1/>)
4. Apache Pig (<http://pig.apache.org/docs/r0.17.0/>)
5. Apache Hive (<https://cwiki.apache.org/confluence/display/Hive/LanguageManual>)
6. Apache Oozie (<https://oozie.apache.org/docs/4.3.1/index.html>)

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Назви розділів, тем	Кількість годин			
	Всього	У тому числі		
		Лекції	Практичні роботи	СРС
Розділ 1. Вступ до Big Data Тема 1.1. Основні поняття Big Data Тема 1.2. Характеристика Big Data Тема 1.3. Приклади задач Big Data Тема 1.4. Сучасні технології Big Data	16	4		12
Розділ 2. Парадигма MapReduce Тема 2.1. MapReduce у розподілених обчисленнях Тема 2.2. Аспекти роботи MapReduce Тема 2.3. Основні етапи MapReduce Тема 2.4. Приклади застосування	24	6	4	14
Розділ 3. Платформа для розподілених обчислень Hadoop Тема 3.1. Характеристики кластеру Hadoop Тема 3.2. Розподілена файлова система – HDFS Тема 3.3. Менеджер ресурсів кластеру – YARN Тема 3.4. Засоби моніторингу роботи кластеру Hadoop Тема 3.5. Запуск обчислювальних задач на кластері Hadoop Тема 3.6. Особливості інсталяція та налаштування кластеру Hadoop	24	6	4	14
Розділ 4. Застосунок Apache Pig Тема 4.1. Основні поняття Apache Pig Тема 4.2. Порівняння з MapReduce Тема 4.3. Мова Pig Latin Тема 4.4. Написання скриптів для кластеру Hadoop	20	6	2	12
Розділ 5. Застосунок Apache Hive Тема 5.1. Основні поняття Apache Hive Тема 5.2. Порівняння між Apache Pig та MapReduce Тема 5.3. Спосіб організації даних у Apache Hive Тема 5.4. Мова запитів HiveQL Тема 5.5. Написання запитів для кластеру Hadoop	18	4	2	12
Розділ 6. Диспетчер робіт Apache Oozie Тема 6.1. Основні поняття Apache Oozie Тема 6.2. Потоки робіт у Apache Oozie Тема 6.3. Особливості налаштування диспетчеру Тема 6.4. Засоби моніторингу робіт Тема 6.3. Створення та налаштування власного потоку робіт для кластеру Hadoop	18	4	2	12
Всього в семестрі	120	30	14	76

Лабораторні заняття:

Метою проведення лабораторних занять є набуття студентами необхідних практичних навичок роботи із сучасними технологіями для організації розподілених обчислень.

- Лабораторна робота №1: Основні поняття MapReduce;
- Лабораторна робота №2: Основні поняття Hadoop;
- Лабораторна робота №3: Основні поняття Apache Pig;
- Лабораторна робота №4: Apache Hive;
- Лабораторна робота №5: Apache Oozie.

6. Самостійна робота студента

- підготовка до лекційних занять по вивченню попереднього лекційного матеріалу;
- підготовка до лабораторних робіт з вивченням теорії лабораторного заняття з усною відповіддю на наведені питання розділу;
- оформлення результатів лабораторної роботи у вигляді протоколу.

Політика та контроль

7. Політика навчальної дисципліни (освітнього компонента)

Під час занять з навчальної дисципліни студенти повинні дотримуватись певних дисциплінарних правил:

- забороняється запізнюватись на заняття;
- не допускаються сторонні розмови або інший шум, що заважає проведенню занять;
- не допускається користування мобільними телефонами та іншими технічними засобами без дозволу викладача.

Лабораторні роботи здаються особисто з попередньою перевіркою теоретичних знань, які необхідні для виконання лабораторної роботи. Перевірка практичних результатів включає перевірку коду та виконання тестових завдань.

Роботи, які здаються із порушенням зазначених термінів без поважних причин, оцінюються на нижчу оцінку.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Поточний контроль: [опитування за темою заняття](#)

Календарний контроль: проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Семестровий контроль: [залік](#)

Умови допуску до семестрового контролю: [зарахування усіх лабораторних робіт](#)

Таблиця 1 — Максимальні бали за окремі лабораторні роботи

Лабораторна	Всього за видом роботи
Лабораторна робота №1	16
Лабораторна робота №2	16
Лабораторна робота №3	16
Лабораторна робота №4	16
Лабораторна робота №5	16
$R_{\text{л}}$	80

Максимальна оцінка за залік (R_z) складає 20 балів:

$$R_z = 20$$

Семестровий рейтинг студента із дисципліни складається з оцінок за: лабораторні роботи (R_l) та залік (R_z).

$$R = R_l + R_z$$

Студент має можливість отримати оцінку за залік автоматом (R_a). Для цього необхідно виконати умови допуску до семестрового контролю до початку залікової сесії. В такому випадку оцінку за дисципліну складатиме:

$$R = R_a = R_l \cdot 1.25$$

В разі якщо роботи здаються у період залікової сесії, студент втрачає право на зарахування заліку автоматом. В цьому випадку максимальний бал за здану в цей період роботу складатиме 60% від зазначеного в таблиці 1.

Таблиця 2 - Відповідність рейтингових балів оцінкам за університетською шкалою:

Кількість балів	Оцінка
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно
Не виконані умови допуску	Не допущено

9. Нормативні положення політики навчальної дисципліни (освітнього компонента)

Правила відвідування занять. Відвідування занять є вільним, бали за присутність на лекціях та практичних заняттях не додаються, відповідно штрафні бали не передбачаються. Втім, вагома частина рейтингу формується через активну участь у заходах на практичних заняттях.

Дистанційний режим навчання. У разі запровадження обмежень на відвідування університету, пов'язаних з введенням карантину або режиму воєнного стану в державі, освітній процес здійснюється у дистанційному режимі відповідно до Положення про дистанційне навчання в КПІ ім. Ігоря Сікорського (<https://osvita.kpi.ua/index.php/node/188>), Регламенту організації освітнього процесу в дистанційному режимі (<https://proffkom.kpi.ua/reglament-organizatsiyi-osvitnogo-protsesu-v-distantsiynomu-rezhimi>) та Регламенту проведення семестрового контролю в дистанційному режимі (<https://osvita.kpi.ua/node/148>). У режимі дистанційного навчання заняття відбуваються у вигляді онлайн-конференції на платформі BigBlueButton. Посилання на онлайн-конференцію та матеріали видається викладачем на початку семестру. Результати оцінювання висвітлюють у АС «Електронний кампус» на особистій сторінці здобувача (<https://ecampus.kpi.ua>).

Правила поведінки на заняттях. На заняттях слід дотримуватись норм етичної поведінки визначених у Кодексі честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>). На території університету здобувачі мають поводити себе відповідно до Правил внутрішнього розпорядку (<https://kpi.ua/admin-rule>).

Визнання результатів навчання, набутих у неформальній/інформальній освіті. Положення про визнання результатів навчання, набутих у неформальній / інформальній освіті (<https://osvita.kpi.ua/index.php/node/179>) регламентує визнання результатів навчання, набутих у неформальній/інформальній освіті:

- у разі навчання на неформальній освіті, яку студент обрав самостійно і хоче зарахувати результати навчання в рамках дисципліни, проходиться процедура валідації, що

передбачає подання здобувачем заяви на ім'я декана, декларації підтверджувальних документів; рішення про визнання чи не визнання приймається комісією у складі завідувача кафедри, викладача, гаранта освітньо-професійної програми.

Політика використання штучного інтелекту. Використання штучного інтелекту регламентується «Політикою використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського» (<https://osvita.kpi.ua/node/1225>). Усі завдання, як під час виконання навчальних завдань з дисципліни, так і індивідуальні завдання, мають бути результатом власної оригінальної роботи здобувача. Використання ШІ для автоматичної генерації відповідей без подальшого їх аналізу та доопрацювання заборонено.

Порушення академічної доброчесності: використання плагіату. У разі виявлення фактів використання плагіату під час виконання навчальних завдань, студент втрачає право на подальше виконання поточних завдань у межах навчальної дисципліни до з'ясування обставин порушення.

Студент зобов'язаний аргументовано пояснити причини використання плагіату, а також засвідчити готовність дотримуватись принципів академічної доброчесності в подальшому навчанні; повторно виконати завдання або, за потреби, виконати додаткове завдання відповідно до бачення викладача з урахуванням вимог щодо самостійності.

Зазначені дії регламентуються Кодексом честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (<https://kpi.ua/code>).

10. Додаткова інформація з дисципліни (освітнього компонента)

Теоретичні та практичні питання, що виносяться під час захисту лабораторних робіт та семестрового контролю, відповідають переліку основних тем, що входять до програми вивчення дисципліни “Технології BigData”.

Робочу програму навчальної дисципліни (силабус):

Складено ст. викладач кафедри обчислювальної техніки, доктор філософії, Таран В. І.

Ухвалено кафедрою обчислювальної техніки (протокол № 12 від 23.06.2025)

Погоджено методичною комісією факультету інформатики та обчислювальної техніки (протокол № 11 від 27.06.2025)